DOCUMENT RESUME

ED 023 195 24 EA 001 727

By-Dingman, Harvey F.; Peck, Robert F.

Reanalysis of Data Files: Dependent Hypotheses and a Recommendation.

Texas Univ., Austin. Research and Development Center for Teacher Education.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-5-0249
Pub Date 68
Contract-OEC-6-10-108
Note-5p.
EDRS Price MF-$0.25 HC-$0.35
Descriptors-*Data Analysis, Information Storage, *Research Problems, *Research Projects, Typology

The reanalysis of data collected in large research projects presents problems, mainly because the hypotheses to be tested may be dependent on the hypotheses of the original project. One possible solution to this problem would be to use the data to group subjects into typologies. (HW)

BR-5-0247
PA-24

REANALYSIS OF DATA FILES: DEPENDENT HYPOTHESES AND A RECOMMENDATION

Harvey F. Dingman, Ph.D.
and
Robert F. Peck, Ph.D.
University of Texas

Abstract

The reanalysis of data collected in large research projects presents problems. It appears that the hypotheses to be tested probably will not be independent. If the data are used to group the subjects into typologies, subsequent research would be improved.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED023195

EA 001 727

REANALYSIS OF DATA FILES: DEPENDENT HYPOTHESES AND A RECOMMENDATION

Harvey F. Dingman, Ph.D.
and
Robert F. Peck, Ph.D.
University of Texas

Large scale research projects often collect valuable pools of data that are not completely analyzed. Much of the data gathered on large projects consist of incidental or background measures that are not related to the major hypotheses being tested. This large pool of data frequently describes a population well enough to be the basis for future studies. At the very least, these data pools provide the normative basis for future experimentation.

While the existence of large data banks has been publicised by organizations (Glaser 1967) the use of the data poses formidable problems. Strictly speaking, if the data file were considered to be a random collection of data representing a population, then the pool could be sampled an infinite number of times with an infinite number of hypotheses. When the probability level to be attached to each hypothesis should be modified by the number of hypotheses tested with those data is not clear. This is particularly ambiguous if the data are sampled by independent investigators with independent hypotheses.

A more practical problem arises when a research project is analyzing a set of data for several purposes. The Austin, Texas, public school system

---

This investigation was sponsored in part by the U.S. Office of Education Contract OE-6-10-108, Research and Development Center for Teacher Education, The University of Texas at Austin.

has been studied for three years by the Research and Development Center at the University of Texas at Austin (Peck, 1968). The teachers who were taught in the experimental teacher education program at the University of Texas have been studied extensively in order to evaluate their training program. A proposal to link the pupils in the schools to the teachers who have received the experimental training requires the interlocking of two data pools for research on some new and previously unplanned comparisons. The hypotheses derived from the combined data pool would not be independent of previously tested hypotheses. Operational use of previously analyzed variables under new names does not increase the information extracted from the data.

The principle that seems apparent is that as long as the new hypotheses and the new comparisons are related operationally to the data that had been derived from the data pool, then the new comparisons are not unbiased. If the number of statistical tests approached is one-half the number of variables and more than fifty per cent of the subjects on each variable, then the only appropriate procedure is to consider the data pool as giving rise to descriptive information which can be used for devising better experiments. The data from the reanalysis of the pool should be especially good for the development of typologies and the evaluation of typological schemes.

The analysis of the file of teachers and pupils should give rise to types of teachers and types of pupils. Efficient new experiments could then be planned to take advantage of the typological scheme in order to maximize the probability of finding significant differences in the new experiments. The replacement of individual variables by typological groupings can enhance the information gained from the experiments as has been shown in the studies with the mentally retarded (Dingman, 1964).

In a series of publications (Dingman, Dingman and Miller, Miller et al. 1961, Miller 1962), a simple typology has been developed and studied. In subsequent papers this typology has been shown to be a formal restatement in mathematical terms of a categorization that has been well recognized (Dingman and Tarjan, Tarj[illegible] et al. 1961). This simple typology, while fairly obvious, can have important new uses since the reorganization of the data and the subjects into objective formal categories (Eyman, in press) can be used in powerful ways. Some of these include the predicting the probability of death and disease for specific patients in an institution (Dingman, et al., 1964). The reuse of the data in a new form affected the careers of mentally retarded patients.

Bibliography

Dingman, H. F. Some uses of descriptive statistics in population analysis. American Journal of Mental Deficiency, 1959, 64, 291-295.

Dingman, H. F A plea for social research in mental retardation. American Journal of Mental Deficiency, (in press).

Dingman, H. F. & Miller, C. R. Note on use of correlation statistics and population analysis. American Journal of Mental Deficiency, 1960, 64, 636-637.

Dingman, H. F. & Tarjan, G. Mental retardation and the normal distribution curve. American Journal of Mental Deficiency, 1960, 64, 991-994.

Dingman, H. F., Tarjan, G., Eyman, R. K., & Miller, C. R. Epidemiology in hospitals: Some uses of data processing in chronic disease institutions. American Journal of Mental Deficiency, 1964, 68, 586-593.

Eyman, R. K. Report of the AAMD ad hoc committee to explore the existence of data files in mental retardation. American Journal of Mental Deficiency, (in press).

Glaser, W. A. Social science data archives in the United States 1967 Council of Social Science Data Archives. 605 West 115 Street, New York, New York, 10025.

Miller, C. R., Eyman, R. K., & Dingman, H F. Factor analysis, latent structure analysis and mental typology. British Journal of Statistical Psychology, 1961, 14, 29-34.

Miller. C. R., Sabagh, G., & Dingman, H. F. Latent class analysis and differential mortality. Journal of the American Statistical Association, 1962, 57, 430-438.

O'Connor, G. & Dingman, H. F. Statistical typologies as a supplement to clinical diagnosis. Behavioral Science, Sept. 1966, 342-345.

Peck, R. F. Annual report of Research and Development Center, Bureau No. 5-0249, April 1, 1968. The University of Texas at Austin, Report No. 11.

Tarjan, G., Wright, S. W., Dingman, H. F. & Eyman, R. K. Natural history of mental deficiency in a state hospital III: Selected characteristics of first admission and their environment. American Journal of Disabled Children, Vol. 101, 1961, 195-205.